

# Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease

Roman Jandarov

Department of Statistics  
The Pennsylvania State University  
raj153@psu.edu

Murali Haran

Department of Statistics  
The Pennsylvania State University  
mharan@stat.psu.edu

Ottar Bjørnstad

Departments of Entomology and Biology  
The Pennsylvania State University  
onb1@psu.edu

Bryan Grenfell

Departments of Ecology and Evolutionary Biology  
Princeton University  
grenfell@princeton.edu

Draft: January 9, 2012

## Abstract

Probabilistic models for infectious disease dynamics are useful for understanding the mechanism underlying the spread of infection. When the likelihood function for these models is expensive to evaluate, traditional likelihood-based inference may be computationally intractable. Furthermore, traditional inference may lead to poor parameter estimates and the fitted model may not capture important biological characteristics of the observed data. We propose a novel approach for resolving these issues that is inspired by recent work in emulation and calibration for complex computer models. Our motivating example is the gravity time series susceptible-infected-recovered (TSIR) model. Our approach focuses on the characteristics of the process that are

of scientific interest. We find a Gaussian process approximation to the gravity model using key summary statistics obtained from model simulations. We demonstrate via simulated examples that the new approach is computationally expedient, provides accurate parameter inference, and results in a good model fit. We apply our method to analyze measles outbreaks in England and Wales between 1944 and 1965. Our method is applicable to problems where traditional likelihood-based inference is computationally intractable or produces a poor model fit. It is also an alternative to approximate Bayesian computation (ABC) when simulations from the model are expensive.

# 1 Introduction

Infectious disease dynamics are of interest to modelers from a range of disciplines. The theory of disease dynamics provides a tractable system for investigating key questions in population and evolutionary biology. Understanding the disease dynamics helps in management and with pressing disease issues such as disease emergence and epidemic control strategies. Probabilistic models for disease dynamics are important as they help increase our understanding of the mechanism underlying the spread of the infection while also accounting for their inherent stochasticity. Observations on reported cases of the diseases, especially in the form of space-time data, are becoming increasingly available, allowing for statistical inference for unknown parameters of these models. However, traditional likelihood-based inference for many disease dynamics models is often challenging because the likelihood function may be expensive to evaluate, making likelihood-based inference computationally intractable. Furthermore, traditional inference may lead to poor parameter estimates and the fitted model may not capture important biological characteristics of the observed data. Hence, an approach that simultaneously addresses the computational challenges as well as the inferential issues would be very useful for a number of interesting and important probabilistic models for dynamics of diseases. Inspired by work in the field of emulation and calibration for complex computer models (cf. Bayarri et al., 2007a; Craig et al., 2001; Kennedy and O’Hagan, 2001; Sacks et al., 1989), we develop a novel approach for inference for such models. Our approach uses a Gaussian process approximation to the disease dynamics model using key biologically relevant summary statistics obtained from simulations of the model at differing parameter values. As we will demonstrate, this approach results in reliable parameter estimates and a good model fit, and is also computationally efficient.

The motivating example for our approach is the gravity time series susceptible-infected-recovered (TSIR) model for measles dynamics. The spatiotemporal dynamics of measles have

received a lot of attention in part due to the importance of the disease, the highly nonlinear outbreak dynamics and also because of the availability of rich data sets. Important aspects of local dynamics of measles are well studied. These include key issues like seasonality in transmission of the infection (Bjørnstad et al., 2002; Dietz, 1976), effects of host demography on outbreak frequency (Finkenstädt et al., 1998; McLean and Anderson, 1988), and causes of local persistence and extinctions (Bartlett, 1956; Grenfell et al., 2001; Grenfell and Harwood, 1997). During the course of outbreaks in well-mixed local populations, the epidemic trajectory of measles is virtually unaffected by infection that may enter from neighboring locations. However, spatial coupling is fundamental to the dynamics and management of measles for smaller communities where the infection may become locally extinct (Bartlett, 1956; Grenfell and Harwood, 1997). Hence, ecologists have also studied the spatial spread of the disease using so-called metapopulation models (Earn et al., 1998; Grenfell and Harwood, 1997; Swinton and Grenfell, 1998).

In this paper, we investigate inference for a model first proposed by Xia et al. (2004). The model represents a combination of the TSIR model (Bjørnstad et al., 2002; Grenfell et al., 2002) with a term that allows for spatial transmission between different host communities modeled as a gravity process. Xia et al. (2004) demonstrate how this model captures scientifically important properties of measles dynamics. Since each likelihood evaluation is computationally very expensive, however, Xia et al. (2004) obtain only point estimates of the parameters minimizing *ad hoc* objective functions instead of using a likelihood-based approach. Here, we develop a more statistically rigorous approach to inferring model parameters, characterizing associated uncertainties and carefully studying parameter identifiability issues. First, in order to explain the issues that arise in inferring these parameters via a likelihood-based approach, we propose a partial discretization of the parameter space that allows us to perform Bayesian inference for the parameters using a fast MCMC algorithm. Using this approach we are able to study uncertainties about the parameter values. The method allows us to investigate parameter identifiability issues, showing which gravity model parameters can or cannot be inferred from a given data set. However, this approach to resolving the computational challenges of traditional likelihood-based inference is problematic, as is revealed by our simulated data examples. We find that the parameter estimates are poor and the forward simulations of the model at these parameter settings do not reproduce epidemiological features of the data deemed key in Xia et al. (2004).

In order to address the above issues, we propose a new approach that directly focuses on the aspects of the underlying process that are of scientific interest. We develop a Gaussian process approximation to the gravity model based on key summary statistics obtained from

simulations of the model at different parameter values. These statistics are chosen by domain experts to capture the biologically important characteristics of the dynamics of the disease. The Gaussian process model ‘emulator’ is then used to develop a probability model for the observations, thereby permitting an efficient MCMC approach to Bayesian inference for the parameters. We demonstrate that the new method recovers the true parameters and the resultant fitted model captures biologically relevant features of the data.

When applied to the gravity TSIR model, our approach allows us to investigate several scientific questions that are of interest to the dynamics of measles. For example, we study changes in dynamics between school holiday periods versus non-holidays. This is particularly interesting because the local, age-structured transmission rate of the disease changes from holidays to non-holidays (Bjørnstad et al., 2002; Dietz, 1976). Since our approach allows us to construct confidence regions easily, we also infer the amounts of exported and imported infected individuals for different cities during different time periods. More generally, the methodology we develop here may be useful for models where the likelihood is expensive to evaluate or in situations where the likelihood is unable to capture characteristics of the model that are of scientific interest. We note that the computational cost of forward simulations for our model makes approaches based on approximate Bayesian computation (ABC) (cf. Beaumont et al., 2002; Marjoram et al., 2003; Pritchard et al., 1999) infeasible. Hence our approach is computationally efficient, while ABC is not a viable option here.

The rest of the paper is organized as follows. Section 2 describes in detail the gravity TSIR model, which acts as our motivating example. Section 3 describes the inferential and computational challenges posed by the model and the large space-time data set. Section 4 describes our new emulation-based approach that is an alternative to traditional likelihood-based inference. Section 5 describes computational details and the application of our method to the gravity TSIR model in simulated data examples. Section 6 describes the application of our method to the England-Wales measles data. Finally, in Section 7, we summarize our results and discuss our statistical approach and scientific conclusions.

## 2 A gravity model for disease dynamics

A general goal of fitting metapopulation disease dynamics models is to describe spatiotemporal patterns of epidemics at the local scale and understand how these patterns are affected by the network of spatial spread of the disease (Cliff et al., 1993; Keeling et al., 2004). The gravity model we study is an extension of a discrete time-series susceptible-infected-recovered model (Bjørnstad et al., 2002; Grenfell et al., 2002) for local disease dynamics which includes

an explicit formulation for the spatial transmission between different host cities (Xia et al., 2004).

The common theoretical framework used to describe the dynamics of infectious diseases is based on the division of the human host population into groups containing susceptible, infected (infectious) and recovered individuals. Let  $I_{kt}$  and  $S_{kt}$  denote the number of infected and susceptible individuals respectively in disease generation  $t$  in city  $k$  and variable  $L_{kt}$  be the number of infected people commuting to city  $k$  at time  $t$ . The ‘commuting’ assumption reflects that movement of infection is mostly through transient movement of individuals. Denote the size and birth rate of city  $k$  at time  $t$  by  $N_{kt}$  and  $B_{kt}$ , and let  $d_{kj}$  represent the distance between cities  $k$  and  $j$ . The model can then be described as follows. First, the model for the number of incidences of measles is

$$I_{k(t+1)} \sim \text{Poisson}(\lambda_{k,t+1}), \text{ where } \lambda_{k,t+1} = \beta_t S_{kt} (I_{kt} + L_{kt})^\alpha, \quad (1)$$

with  $t = 1, \dots, T, k = 1, \dots, K$ . The time-step is taken to be 2 weeks, roughly corresponding to the generation length (serial interval) of measles. The so-called transmission coefficient,  $\beta := \{\beta_t\}$ , is a parameter that represents the attack rate of measles at time  $t$  and  $\alpha$  is a positive real number correcting for the discrete-time approximation to the underlying continuous-time epidemic process (Glass et al., 2003). Since these parameters only affect the local dynamics of measles, henceforth we refer to these parameters as local dynamics parameters. The indexing by  $t$  for  $\beta_t$  reflects how this parameter is taken to be a piece-wise constant taking 26 different values to accommodate seasonal variability of the transmission rate that is repeated every year (Bjørnstad et al., 2002; Fine and Clarkson, 1982; Finkenstädt and Grenfell, 2000; Grenfell et al., 2002). From this, it can be seen that  $I_{k(t+1)}$  increases depending on the number of susceptibles and the number of moving infections coming to city  $k$  at the previous time step. Note that we use the Poisson distribution whereas Xia et al. (2004) use the Negative Binomial distribution; this is due to the greater computational stability of the Poisson distribution for small values of  $\lambda$ . Our exploratory analysis show that a model fit from using the Poisson distribution is similar to a model fit obtained with the Negative Binomial distribution.

The susceptibles are modeled as follows

$$S_{k(t+1)} = S_{kt} + B_{kt} - I_{k(t+1)}, \quad (2)$$

reflecting how susceptibles are replenished by births and depleted by infection. Since case fatality from measles was very low for the period of time in this study and mean age of infection was small, mortalities are not included in this balance equation.

Finally, the gravity model describes the number of moving infected individuals by

$$L_{kt} \sim \text{Gamma}(m_{kt}, 1), \text{ where } m_{kt} = \theta N_{kt}^{\tau_1} \sum_{j=1, j \neq k}^K \frac{I_{jt}^{\tau_2}}{d_{kj}^{\rho}}, \quad (3)$$

where  $\text{Gamma}(a, b)$  represents the Gamma distribution with shape and scale parameters  $a$  and  $b$  respectively. The reason to model immigrant infection as a continuous random variable lies in the assumption that the transient infectives do not remain for a full epidemic generation.

The local dynamics parameters in Equation (1) have been estimated previously (Bjørnstad et al., 2002; Finkenstädt et al., 2002; Grenfell et al., 2002). In this study, we are interested in learning about the parameters  $\theta$ ,  $\tau_1$ ,  $\tau_2$  and  $\rho$  in Equation (3) as these parameters control the spatial spread and regional behavior of the disease. Note, however, that for convenience and numerical stability, we use a reparametrization of  $\theta$ ,  $\theta' = -\log_{10}(\theta)/5$  throughout the paper.

### 3 Parameter inference for the gravity model

Reliable estimates of the local dynamics parameters  $\alpha$  and  $\beta$  are available for measles dynamics (Bjørnstad et al., 2002; Finkenstädt et al., 2002; Grenfell et al., 2002; Xia et al., 2004). Therefore, since we are interested in spatial dynamics here, we assume that these parameters are known and use the previously obtained estimates as the true values. In particular, the local seasonal transmission parameters for biweeks 1 through 26,  $\beta_t$ , is taken to be equal to  $\beta_t = (1.24, 1.14, 1.16, 1.31, 1.24, 1.12, 1.06, 1.02, 0.94, 0.98, 1.06, 1.08, 0.96, 0.92, 0.92, 0.86, 0.76, 0.63, 0.62, 0.83, 1.13, 1.20, 1.11, 1.02, 1.04, 1.08)$ , and  $\alpha$  is assumed to be 0.97. Given parameter identifiability issues, joint estimation of the spatial dynamics and all the local dynamics parameters ( $\alpha$  and  $\beta_t$ ) is infeasible. Furthermore, assuming that the local dynamics parameters are known does not have an undue effect on the model fit as was shown in the literature (cf. Xia et al., 2004). This leaves us with four unknown parameters,  $\theta'$ ,  $\tau_1$ ,  $\tau_2$  and  $\rho$ , that we call the gravity model parameters (in our Gaussian process based approach in Section 4 we will also introduce several other parameters). In this paper our focus is on investigating the gravity model parameters and, when possible, obtaining the best estimates of them with relevant descriptions of their variability.

As suggested by our domain experts, feasible values for the gravity parameters lie in the interval  $[0, 2]$  (see also Xia et al., 2004). Therefore, we use uniform priors for  $(\theta', \tau_1, \tau_2, \rho)$  in all the inferential approaches that follow.

The data are spatiotemporal and tend to be high-dimensional,  $546 \times 952$  in the case of the England-Wales measles data. To study whether our fitted model captures epidemiologically relevant features of the data, we focus on two important biological characteristics of the process as suggested by domain experts. These are:

1. Maximum number of incidences which we will denote by  $\mathbf{M} = (M_1, \dots, M_K)$ , where  $M_i$  is the maximum number of incidences for the  $i$ -th city.
2. Proportions of bi-weeks without any cases of infection denoted by  $\mathbf{P} = (P_1, \dots, P_K)$ , where  $P_i$  is the proportion of incidence free biweeks for the  $i$ -th city.

An important goal of our work is to find parameter settings (along with associated uncertainties and dependencies among them) that yield a model that produces disease dynamics that are as close as possible to the data in terms of capturing these key properties.

### 3.1 A gridded MCMC approach

In this section we demonstrate via simulated data examples how traditional likelihood-based approaches can be problematic for the gravity model. Because likelihood-based inference is computationally intractable, we develop a gridded MCMC approach instead; this approach involves a partial discretization of the parameter space. The gridded MCMC algorithm also requires certain simplifying assumptions and data imputation for unobservable susceptibles ( $\{S_{kt}\}$ ) which are explained below. We would like to clarify, however, that the alternative inferential method we later recommend in Section 4 neither requires data imputation nor any of the simplifying assumptions used for the gridded MCMC approach.

It is easy to see why each evaluation of the likelihood for the gravity model is expensive. As in many population dynamic models, the major difficulty is in integrating over high-dimensional unobserved variables. For our model,  $\{L_{kt}\}$  and  $\{S_{kt}\}$  are of  $K \times T$  dimensions each, which translates to  $2 \times 519,792$  in the case of measles data set considered in Section 6. Details of the likelihood function are given in Appendix A.

In order to avoid integrating over the latent  $\{S_{kt}\}$ , we first use a standard susceptible reconstruction algorithm (Bobashev et al., 2000; Ellner et al., 1998; Fine and Clarkson, 1982; Finkenstädt and Grenfell, 2000; Schenzle, 1984) that is based on the fact that in the pre-vaccination era almost all children were infected. This algorithm allows us to impute  $\{S_{kt}\}$  from the observations for  $\{I_{kt}\}$ ,  $\{N_{kt}\}$  and  $\{B_{kt}\}$  by the following. We rewrite Equation (2)

of the gravity model according to Finkenstädt and Grenfell (2000):

$$S_{kt} = \overline{S}_k + D_{k0} + \sum_{j=0}^t B_{kj} - \sum_{j=0}^t I_{kj}/\omega \quad (4)$$

where  $\overline{S}_k$  is the mean number of susceptibles in city  $k$ ,  $D_{k0}$  is the unknown deviations around the mean at time 0, and  $\omega$  is the reporting rate. We take this rate to be equal to 0.5 which is a common assumption for measles dynamics in the prevaccination era based on several other studies (cf. Bjørnstad et al., 2002; Clarkson and Fine, 1985; Finkenstädt and Grenfell, 2000). In practice, there are variations in reporting rates with infection level for different locations. This does not, however, seem to be an issue for susceptible reconstruction for measles dynamics (cf. Clarkson and Fine, 1985).

We can reconstruct the time series  $D_{kt}$  of how the local susceptible numbers deviate from the local mean value,  $D_{kt} = S_{kt} - \overline{S}_k$ , by rewriting Equation (4) as,

$$\sum_{j=0}^t B_{kj} = D_{k0} + 1/\omega \sum_{j=0}^t I_{kj} + D_{kt} \quad (5)$$

from which it is clear that  $D_{kt}$  is the residual from the regression of cumulative number of births on the cumulative number of cases. Note here that this algorithm works when  $D_{k0}$  and the reporting rate  $\omega$  are unknown since these are accommodated by the intercept and slope of the cumulative-cumulative regression. The method, however, does not allow the independent estimation of the mean number of susceptibles. Using the previous analysis, we assume that this number is equal to 4% of the population in city  $k$  (Bjørnstad et al., 2002).

We avoid integrating over the unobserved  $\{L_{kt}\}$  by setting  $L_{kt} = m_{kt}$  for all  $k$  and  $t$ , that is, using the expectation instead of using the full Gamma distribution. Based on a study of this in several simulated examples (where we know the true values of  $\{L_{kt}\}$ ), this assumption does not seem to affect our likelihood-based inference and conclusions.

As expected, after these steps of simplification and data imputation, likelihood calculations are faster, but still expensive taking more than two minutes to evaluate for each parameter setting. This computational efficiency is sufficient if we want to make inference based only on maximization of the likelihood. It takes about 72 hours for a standard optimizer in R (R Development Core Team, 2011) to converge on an AMD Quad Core 2.6 GHz processor. However, a much more thorough exploration of inference for this model is desirable; we are interested in learning about parameter uncertainties, the joint distributions of the parameters, as well as any identifiability issues. We achieve the additional speedup necessary for MCMC by using a gridded MCMC approach as follows. Note that for each



calculation of the likelihood, we calculate  $K \times T$  values of  $m_{kt}$  according to Equation 3. For each  $k$  and  $t$ , this requires summing over  $K$  quantities. We speed our calculations by selecting a grid on the range of possible values for  $\tau_2$  and  $\rho$ . For each point of the grid, we then calculate and save a set of computationally expensive matrices  $\{M_{kt}(\tau_2, \rho)\}$ :

$$M_{kt}(\tau_2, \rho) = \sum_{j=1, j \neq k}^K \frac{(Ijt)^{\tau_2}}{d_{kj}^\rho}.$$

Then, when evaluating the likelihoods,  $L(\theta', \tau_1, \tau_2, \rho)$ , we use the pre-calculated matrices  $\{M_{kt}(\tau_2, \rho)\}$  and the fact that

$$m_{kt} = \theta' N_{kt}^{\tau_1} M_{kt}(\tau_2, \rho).$$

In this way, we reduce the number of arithmetic operations necessary to calculate the likelihood for each iteration from  $O(K^2T)$  to  $O(T + K)$ . For the measles data we analyze in Section 6.2, this reduces the number of floating point operations involved in each likelihood evaluation from 499,009,056 to 1,502.

With discretized  $\tau_2$  and  $\rho$ , we make inference based on the posterior distribution of the parameters using samples obtained via MCMC. Details of inference based on the approach and relevant plots of the inferred posterior surface are summarized in Sections 3.2 and 5.1.

### 3.2 Simulated examples

We note that all simulated data sets we consider in this work are generated from the full gravity model described in Section 2 with initial points equal to the actual observations at  $t = 1$ . In these examples, the number of locations, their coordinates, demographic variables, and the number of time steps are the same as those in the measles data described in Section 6.1.

In our first example, we simulate a data set using values for the gravity parameters  $\theta' = 0.71$ ,  $\tau_1 = 0.3$ ,  $\tau_2 = 0.7$  and  $\rho = 1$ . This parameter setting results in realistic data that resembles the observations. Figure 1 shows conditional and unconditional posterior likelihood surface plots for  $\theta'$  and  $\rho$  obtained by using the above gridded MCMC approach. From these plots, we can easily see that inference for  $\theta'$  and  $\rho$  is not possible because of the apparent issue with identifiability (Figure 1 (a)). In Figure 1 (b) we see that identifiability is reduced, but still exists when we fix one of the parameters, say  $\tau_1$ , at its known true value. In Figure 1 (c), we fix both of  $\tau_1$  and  $\tau_2$  at their true values and see that the obtained ridge contains the true values for  $\theta'$  and  $\rho$ . Figure 1 (d) demonstrates that the ridge moves by changing the values of  $\tau_1$  and  $\tau_2$  away from their true values. Figure 2 is similar to Figure 1. The

difference is that here we plot everything in  $\tau_2$  and  $\rho$  surface first without any assumptions (Figure 2 (a)), then by fixing  $\theta'$  (Figure 2 (b)) and then by fixing both  $\theta'$  and  $\tau_1$  (Figure 2 (c)). Finally, Figure 2 (d) shows how the ridge in the posterior moves by changing the values of the other parameters.

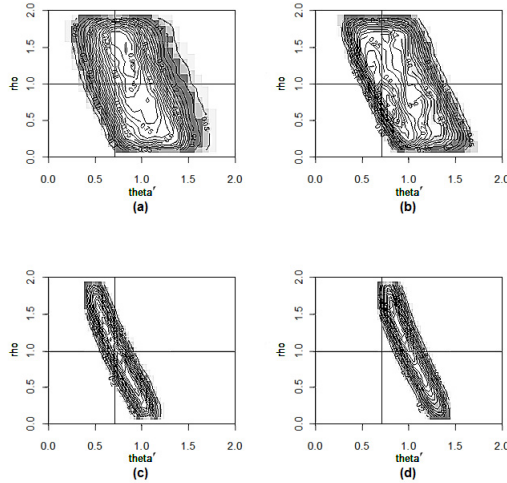


Figure 1: Inferred posterior 2D likelihood surface obtained for data with known parameters ( $\theta' = 0.71$ ,  $\tau_1 = 0.3$ ,  $\tau_2 = 0.7$  and  $\rho = 1$ ): (a) Marginal 2D likelihood surface for  $(\theta', \rho)$ ; (b) Marginal 2D likelihood surface for  $(\theta', \rho)$  assuming  $\tau_1 = 0.3$  (true); (c) 2D likelihood surface for  $(\theta', \rho)$  assuming  $\tau_1 = 0.3$  (true) and  $\tau_2 = 0.7$  (true); (d) 2D likelihood surface for  $(\theta', \rho)$  assuming  $\tau_1 = 0.5$  (any value) and  $\tau_2 = 1$  (any value).

In our second example, we simulate a data set using values for the gravity parameters  $\theta' = 0.71$ ,  $\tau_1 = 0.5$ ,  $\tau_2 = 1$  and  $\rho = 1$ . Figure 3 is a plot of the two-dimensional likelihood in  $\theta'$  and  $\rho$  space obtained by fixing  $\tau_1$  and  $\tau_2$  at their true values 0.5 and 1 respectively. We can see here that the true values of the parameters of interest are not in the region where the likelihood is maximized. This, unfortunately, means that repeating the above with other simulated data with different true values for the gravity parameters reveals that the ridge analogous to the ridge in Figure 1 (c) does not always have to contain the true values for  $\theta'$  and  $\rho$ . From our study of multiple simulated data, we also find that the likelihood ridge can have an intercept that is different from the ridge that we would intuitively think as the true ridge while having the same slope. This difference in intercepts creates a shift thereby resulting in poor parameter inference. Unfortunately the magnitude and direction of the shift depends on the true parameter values, so no simple bias correction is available. At first, one may think that the discretization of the parameters  $\tau_2$  and  $\rho$  may be causing some of

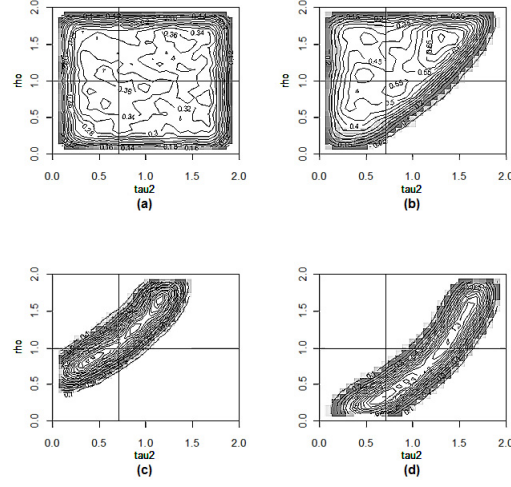


Figure 2: Inferred posterior 2D likelihood surface obtained for data with known parameters ( $\theta' = 0.71$ ,  $\tau_1 = 0.3$ ,  $\tau_2 = 0.7$  and  $\rho = 1$ ): (a) Marginal 2D likelihood surface for  $(\tau_2, \rho)$ ; (b) Marginal 2D likelihood surface for  $(\tau_2, \rho)$  assuming  $\theta' = 0.71$  (true); (c) 2D likelihood surface for  $(\tau_2, \rho)$  assuming  $\theta' = 0.71$  (true) and  $\tau_1 = 0.3$  (true); (d) 2D likelihood surface for  $(\tau_2, \rho)$  assuming  $\theta' = 1$  (any value) and  $\tau_1 = 0.3$  (any value).

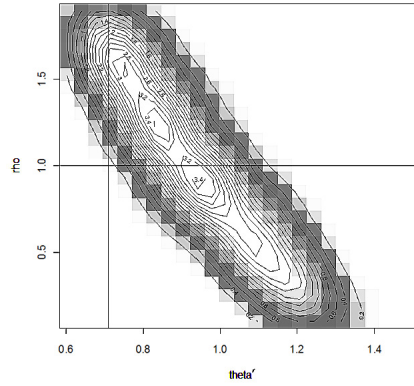


Figure 3: Inferred posterior 2D likelihood surface obtained for data with known parameters ( $\theta' = 0.71$ ,  $\tau_1 = 0.5$ ,  $\tau_2 = 1$  and  $\rho = 1$ ): Posterior 2D likelihood surface for  $(\theta', \rho)$  assuming  $\tau_1 = 0.5$  (true) and  $\tau_2 = 1$  (true) has a shift and does not contain the true  $(\theta', \rho)$  at its highest probability area.

these issues. We verify that this is not the case by simply computing the values of the true likelihood function at the top of the ridges obtained with the discretized likelihood. We are able to see that the likelihood surface using the discretization is similar to the true likelihood surface. The poor inference from our traditional Bayes approach is therefore clearly not a result of the discretization.

By generating additional simulations using a simpler model where we fix all the latent variables at their means we also find the full gravity model does not substantially differ from the simpler one in terms of capturing interesting biological characteristics of the underlying dynamics of the disease. In order to study the effect of this fixing on the likelihood surface, we save the true latent variables while simulating data and use them in our gridded MCMC in place of the expectations used in our gridded MCMC algorithm. The results show that using the true values of the latent variables does not change the traditional Bayes inference. This also confirms that the shifts that we observe in the traditional Bayes approach are not due to simplifying the model, but rather due to inherent problems with the likelihood function.

We note that our main interest is to examine whether the parameter estimates result in a model fit that is capable of reproducing important characteristics of the observations. In order to study the model fit from the gridded MCMC, we simulate a data set using the full gravity model with estimated values of the parameters, where here and throughout the paper, we use modes of the corresponding posterior density functions as estimates of the parameters. These estimates for the measles data described in Section 6.1 are  $(\theta', \tau_1, \tau_2, \rho) = (0.71, 0.5, 1, 1.48)$ . For the simulated data set, we calculate the two 952 dimensional vectors (number of cities in the data) of summary characteristics and plot them against the summary vectors for the observed measles data (Figure 4). We can see that the simulated data do not seem to match the actual data in terms of the maximums  $\mathbf{M}$  and the proportions of zeros  $\mathbf{P}$  (Figure 4 (a)-(b)). In Section 5.2, we compare the model fit obtained via the gridded MCMC to the model fit we obtain via our Gaussian process-based approach described in Section 4.

We summarize below our conclusions based on the gridded MCMC approach:

1. The confidence regions for the parameters are very wide, suggesting that there may be relatively little information even with a fairly rich data set. Hence we assume that  $\tau_1 = 1$ ,  $\tau_2 = 1$  as estimated in Xia et al. (2004) and study the joint distribution of  $\theta'$  and  $\rho$ , which becomes well informed by the data.
2. The fitted gravity model, using the above inference about its parameters, does not capture important biological features of the data.
3. We find that the parameter estimates from the traditional Bayes approach are shifted

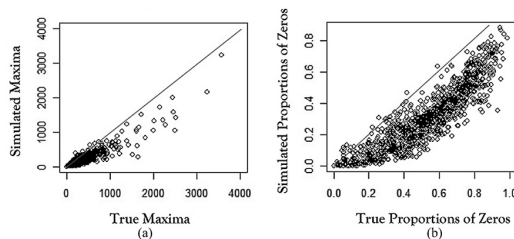


Figure 4: Characteristics of simulated data at the parameters obtained via the traditional Bayes approach: (a) Simulated  $\mathbf{M}$  vs  $\mathbf{M}$  from the data; (b) Simulated  $\mathbf{P}$  vs  $\mathbf{P}$  from the data.

and the direction of the shift varies as shown in Figure 3. For example, for a simulated data set using the parameters values  $(\theta', \tau_1, \tau_2, \rho) = (0.71, 0.5, 1, 1)$ , our attempt to infer  $\rho$  assuming other parameters are known results in an estimate  $\hat{\rho} = 1.5$  with a confidence region that does not contain the truth .

## 4 Gaussian processes for emulation-based inference

Since a traditional Bayes approach suffers from the above shortcomings, we develop an alternative method that is directly linked to the characteristics of the infectious disease dynamics that are of most interest to biologists.

We describe a new two-stage approach for inferring the gravity parameters. In the first stage, we simulate the gravity model at several parameter settings. For each forward simulation of the model we can calculate the vector of summary statistics based on the simulated data set. This vector is high-dimensional, 952 dimensions in the case of measles data. Since Gaussian process-based emulation for high dimensions poses serious computational challenges, we emulate the model by fitting a Gaussian process to the Euclidean distances between the summary statistics of the simulated data at the chosen parameter settings and the summary statistics for the real data. In the second stage, we perform Bayesian inference for the observations using the GP emulator from the first stage. We also allow for additional sources of uncertainty such as observational error and model-data discrepancy as described below. We note that such two-stage approaches to parameter inference in complex models has been used to reduce computational challenges and alleviate identifiability issues (cf. Bhat et al., 2011; Liu et al., 2009).

We begin with some notation. Let  $Z$  denote the vector of summary statistics of interest

(e.g. proportions of zeros) calculated using the observed space-time data set. Let  $\Theta$  be the gravity parameters and  $Y(\Theta)$  denote the vector of summary statistics obtained using a simulation from the gravity model with the parameter setting  $\Theta$ . Let  $\Omega = (\Theta_1, \dots, \Theta_p)$  be a grid on the parameter space. Our first goal is then to model  $\mathbf{D} = (D_1, \dots, D_p)$ , where  $D_i$  is the Euclidean distance between  $Y(\Theta_i)$  and  $Z$  for  $i = 1, \dots, p$ . This is done in the first stage of our approach where we assume,

$$\mathbf{D}|\Omega, \beta_{\mathbf{G}}, \xi_{\mathbf{G}} \sim N(X\beta_{\mathbf{G}}, \Sigma(\xi_{\mathbf{G}})) \quad (6)$$

where  $X$  is a design matrix of dimension  $p \times 5$  with  $i$ -th row equal to  $(1, \Theta_i^T)$ . In other words, columns of  $X$  are the values the gravity parameters,  $(\theta, \tau_1, \tau_2, \rho)$ , on the selected grid and an intercept. We use Gaussian covariance matrix,  $\Sigma(\xi_{\mathbf{G}})$ , elements of which are given by,

$$\begin{aligned} (\Sigma(\xi_{\mathbf{G}}))_{ij} &= \text{cov}(D_i, D_j) = \\ &= \begin{cases} \sigma_G^2 \exp(-\phi_G^2 \|\Theta_i - \Theta_j\|^2), & \text{if } i \neq j \\ \sigma_G^2 + \tau_G^2, & \text{otherwise.} \end{cases} \end{aligned}$$

Here,  $\|a - b\| := d(a - b, a - b)$ , where throughout the paper, the function  $d(\cdot, \cdot)$  returns the Euclidean distance between the argument vectors.  $\xi_{\mathbf{G}} = (\sigma_G^2, \tau_G^2, \phi_G)$  is a vector of parameters that specify the covariance matrix, and  $\beta_{\mathbf{G}}$  is a vector of regression coefficients. Then, if we let the maximum likelihood estimate of  $(\beta_{\mathbf{G}}, \xi_{\mathbf{G}})$  be  $(\hat{\beta}_G, \hat{\xi}_G)$ , using standard multivariate normal theory (cf. Anderson, 1984), the normal predictive distribution for the simulated distance  $D$  at a new  $\Theta$  can be obtained by substituting  $(\hat{\beta}_G, \hat{\xi}_G)$  in place of  $(\beta_G, \xi_G)$  and conditioning on  $\mathbf{D}$ . We denote this predictive distribution by  $\eta(D; \Theta)$ . Detailed version of constructing this predictive distribution (emulator) is given in Appendix B.

Consider a new space-time data set, and let the vector of summary statistics for these data be  $Y^*$ . Let the distance between  $Y^*$  and  $Z$  be  $D^*$ . The predictive distribution from the first stage provides a model for  $D^*$ ,  $\eta(D^*; \Theta^*)$ , connecting it to some unknown parameter vector  $\Theta^*$ .

Following Bayarri et al. (2007b), we model the discrepancy between the gravity model and the real data. Failing to account for data-model discrepancy can lead to poor inference as pointed out in Bayarri et al. (2007b) and Bhat et al. (2010). We account for this by setting  $D^* = D_{\delta}^* := \delta$ , where  $\delta > 0$  is the discrepancy term. It is positive since it represents an Euclidean distance that is non-negative (in the unrealistic case that there is an exact match between the model for the data and the model used to fit the data,  $\delta$  would be identically equal to 0). We then infer the gravity parameters using  $\eta(D_{\delta}^*; \Theta^*)$  considering  $\delta$  to be another unknown parameter in the MCMC algorithm. In other words, the likelihood

function we use for our MCMC algorithm is a function  $f(\delta, \Theta^*) := \eta(D_\delta^*; \Theta^*)$ . We note that including a model discrepancy term results in more reliable parameter inference with narrower confidence regions since it adjusts for the fact that even the best model fit is not going to reduce the distance between the simulated and observed summary statistics to zero. In our simulated examples, where data are generated from the gravity model, the discrepancy term can be thought of as an adjustment parameter for the fact that two data sets simulated at the same parameter settings will always have small differences due to stochasticity. In these examples, as it is expected, estimate of the discrepancy is very small compared to the discrepancy term inferred from the original data. We also note that using negative values for  $\delta$  would mean an extrapolation in our emulator beyond the grid of the parameter space that may lead to unreliable inference. In many situations, having a well-defined discrepancy term with an informative prior helps to reduce problems with identifiability of the parameters as well (cf. Craig et al., 2001).

We can now summarize our inferential approach as follows:

1. Emulating the gravity model:
  - (a) Select a grid  $(\Theta_1, \dots, \Theta_p)$  on the range of possible values for  $\Theta$ .
  - (b) Calculate  $Y(\Theta_i)$  using a simulation from the gravity model with  $\Theta_i$  for all  $i$ .
  - (c) Calculate  $\mathbf{D} = (D_1, \dots, D_p)$ , distances from  $Y_i$  to  $Z$  for all  $i$ .
  - (d) Find the maximum likelihood estimates of  $(\beta_{\mathbf{G}}, \xi_{\mathbf{G}})$ , the parameters of the Gaussian process in Equation (6). Obtain the predictive distribution  $\eta(D; \Theta)$ .
2. Bayesian inference for  $\delta$  and  $\Theta^*$  given the observations  $Z$ :
  - (a) Using the predictive distribution with a discrepancy term,  $\eta(D_\delta^*; \Theta^*)$ , perform Bayesian inference for the parameters  $(\Theta^*, \delta)$  from the posterior distribution via MCMC.

## 5 Emulation-based inference for the gravity TSIR model

In this section we describe details of the application of the inferential approach described in Section 4 to the gravity TSIR model. By using simulated data examples, we show that the approach resolves the problems posed by traditional approaches. In order to contrast our approach to a traditional likelihood-based approach (carried out by gridded MCMC as described in Section 3.1), we also provide computational details from the application of both methods.

## 5.1 Computational details of gridded MCMC and emulation-based approaches

Inference for both the traditional Bayes and emulator-based approaches relies on sampling from the corresponding posterior distributions via MCMC. In both methods, we use univariate sequential slice sampling updates for the continuous parameters (Agarwal and Gelfand, 2005; Neal, 2003). Parameters that are on the grid are updated via an analog of a simple random walk for discrete variables. In all the MCMC algorithms that are used for the discretized MCMC approach, the chain is run until we obtain 200,000 samples. This takes about 3 days on a Intel Xeon E5472 Quad-Core 3.0 GHz processor. In all the MCMC algorithms for the Gaussian process-based method, all the updates are carried out using slice sampling since all the parameters here are continuous. Chain lengths are 200,000 again and it takes about 10 hours to generate them. The chain lengths in both methods are adequate for producing posterior estimates with small Monte Carlo standard errors (Flegal et al., 2008; Jones et al., 2006).

We emulate the gravity model with a Gaussian process using proportions of zeros as a summary statistic of interest. Using different summary statistics may, of course, lead to different inference. Inference based on the maximums, however, was identical to what is obtained here and therefore we do not include details of the analysis and the corresponding results. It is also possible to develop an emulator using these two summary statistics at the same time; this is computationally more demanding and based on our exploratory data analysis, it will not impact our conclusions. In general, summary statistics for emulation need to be chosen to capture the most scientifically important aspects of the disease dynamics process.

We use the priors for the gravity model parameters that are described in Section 3. Since the discrepancy term,  $\delta$ , is always positive, we use an exponential(1) as its prior distribution.

We use a uniform grid in the four-dimensional cube, each side of which is equal to the intervals  $[0, 2]$ . For each parameter, we use 20 different values on each axis of the cube; this grid size permits computationally expedient inference. Our analysis of simulated data sets also shows that 20 is sufficient for accurate inference. In addition, for each point on the grid, the average distances from multiple forward simulations can be used instead of the distances calculated from a single simulation. This may be important when model realizations are highly variable. For the parameters of the gravity model, however, our inference was insensitive to the number of repetitions as the model had relatively small amount of stochasticity.



## 5.2 Application to simulated data

In the simulated examples that follow, our goal is to compare inference based on the GP-approach to inference from the traditional Bayes approach. In Figure 5, we show a simulated example where both the GP and traditional Bayes approaches yield the same inference, and another simulated example where the two approaches yield different answers. In both cases, the emulation-based approach provides inference that captures the true parameter values. In the first simulated data, the true parameters are  $\theta' = 1$ ,  $\tau_1 = 0.6$ ,  $\tau_2 = 1$  and  $\rho = 1$ . In Figure 5 (a), we overlay two different 95% confidence regions obtained using the two different methods. Both of these regions are found by assuming  $\tau_1 = 0.6$  and  $\tau_2 = 1$ . We can see that for this example, both solid (traditional Bayes) and dashed (GP emulator-based) regions contain the true values of  $\theta'$  and  $\rho$ . This shows that inference based on the GP emulator is as good as inference based on the traditional Bayes method. To demonstrate that the new approach is better than the traditional Bayes approach, we choose a second set of values for the gravity parameters ( $\theta' = 0.71$ ,  $\tau_1 = 0.62$ ,  $\tau_2 = 1$  and  $\rho = 1.5$ ) for which we know inference based on the traditional Bayes approach to be poor (like in Figure 3). Figure 5 (b) shows how the 95% confidence region from the traditional Bayes method (outlined with a solid line) is shifted and does not contain the truth. The permissible region obtained using the GP emulator (outlined with a dashed line) have corrected the shift and contains the true values of the parameters.

We analyze the ability of the fitted gravity model to reproduce the key characteristics of the process at these new parameter estimates. Using estimates obtained via the GP-emulator based approach,  $(\theta', \tau_1, \tau_2, \rho) = (0.71, 0.5, 0.5, 1.48)$ , we generate a data set to obtain plots similar to the ones in Figure 4. Plots on Figure 6 (a)-(b) show that the model now can fit the maximums  $\mathbf{M}$  and the proportions of zeros  $\mathbf{P}$  very well. Comparing the plots in Figures 4 and 6, we can now say that the new emulation-based approach improves the model fit substantially while the traditional Bayes parameter estimates from the gridded MCMC fail to provide a model that captures the key epidemiological features of the data.

In order to study the effect of a discrepancy term in our approach, we also tried to infer the gravity parameters using the emulation-based model with  $\delta = 0$  (no discrepancy). The resultant 95% confidence regions were much wider for the latter approach containing incorrect parameter settings, supporting the points made in Bayarri et al. (2007b) about the importance of adding a discrepancy term to approximate models. We also tried a few different priors; using the exponential(1) prior for the discrepancy term worked very well as was clear from the results: the posterior median for the discrepancy term was found to be around 2 which was close to the minimal distance from the simulated and the true vectors

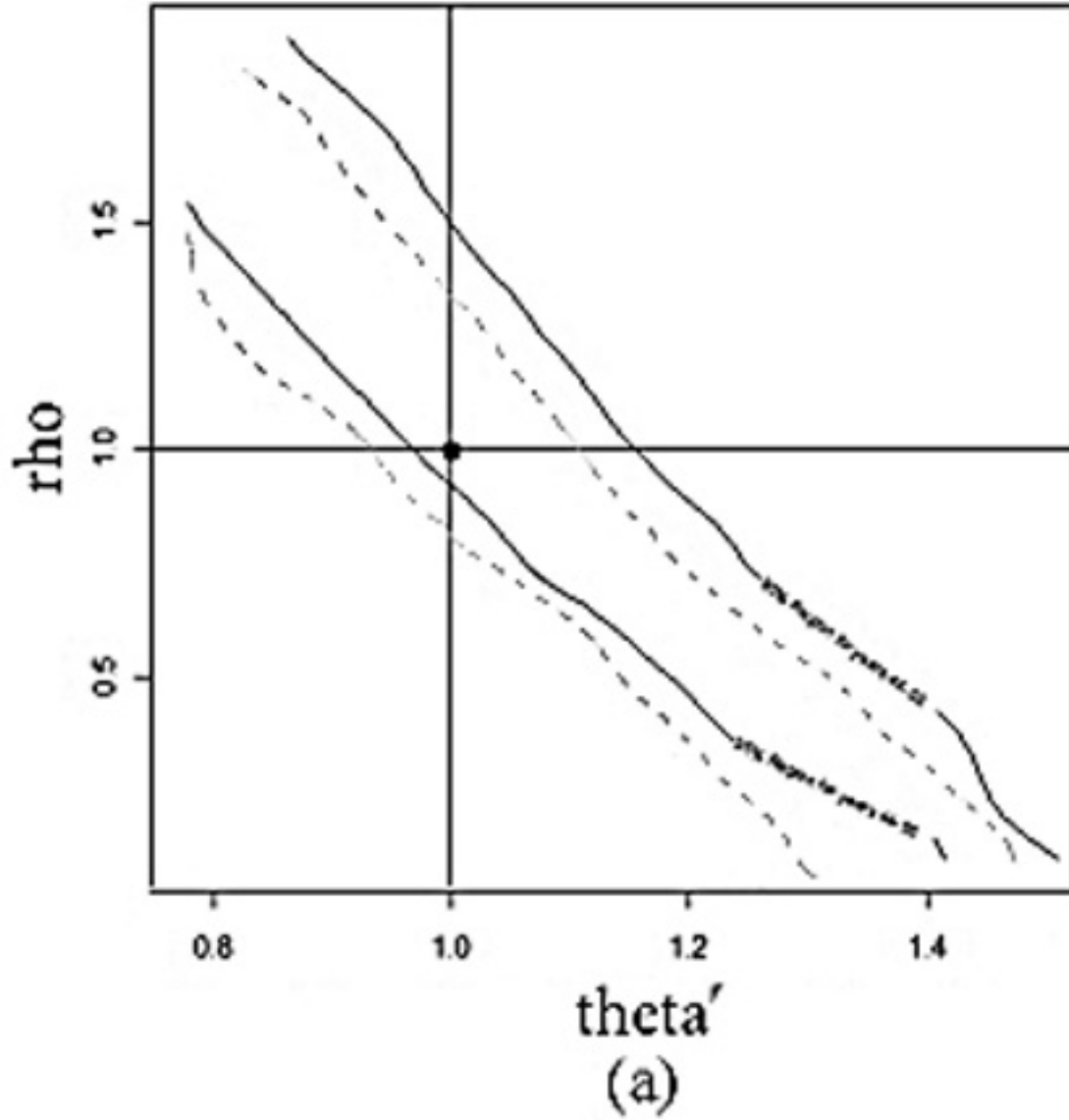


Figure 5: 95% C.I.'s for  $(\theta', \rho)$  obtained via different methods (assuming that  $\tau_1$  and  $\tau_2$  are known): Solid line shows the 95% region obtained using the traditional Bayes method. Dashed line outlines the 95% region obtained via GP emulator: (a) Both regions contain the true parameter values; (b) Region obtained by the GP emulator contains the true values of the parameters, while the traditional Bayes region does not.

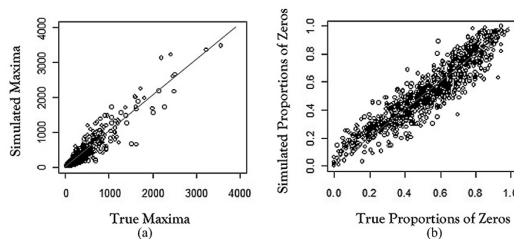


Figure 6: Characteristics of simulated data at the parameters chosen to minimize the discrepancy between the data and the simulation: (a) Simulated  $\mathbf{M}$  vs  $\mathbf{M}$  from the data; (b) Simulated  $\mathbf{P}$  vs  $\mathbf{P}$  from the data.

of summary statistics taken over all the points on the grid.

## 6 Results from application to measles data

We apply our emulation-based approach to inference for the gravity TSIR model to a well known measles data set from the U.K. The purpose of this is twofold: to demonstrate the applicability of our approach to a real data set as well as to provide some insights into measles dynamics in the pre-vaccination era.

### 6.1 Description of measles data set

The following description of the data closely follows Xia et al. (2004). We analyze weekly case reports of measles for the 952 locations in England and Wales from 1944 to 1967. The data represent an interesting case study of spatiotemporal epidemic dynamics (Grenfell et al., 2002) with well understood underreporting rate of 40%-55% (Bjørnstad et al., 2002). Besides the under-reporting, the data are complete and reveal inter-annual outbreaks of infection. A critical feature of this data set is that, except for a few large cities, infection frequently goes locally extinct, so that overall persistence hinges on episodic reintroduction and spatial coupling. Before further analysis, we correct the reported data by a factor of  $1/0.52$ , with 52% being the average reporting rate taken from previous analysis (Bjørnstad et al., 2002; Clarkson and Fine, 1985; Finkenstädt and Grenfell, 2000). In addition, as in previous works, we use a timescale that represent the exposed and infectious period, which is known to be about 2 weeks for measles (Black, 1989).

Following a standard assumption in the literature (see, for instance, Bjørnstad et al., 2002;

Grenfell et al., 2002; Xia et al., 2004, and the references therein), the population sizes and per capita birth rates for all locations in this work are assumed to be approximately constant throughout the time period. These variables are taken as those in 1960 for each of the areas. This is a rough approximation, since most communities grew during the period we analyze. The force of infection is, therefore, on average slightly underestimated (overestimated) during the early (late) part of the study.

## 6.2 Some implications for measles dynamics

Important biological questions we want to answer based on these data are: (i) do the gravity model parameters (and hence disease transmission) change for different time periods? Do they change for school holiday periods versus non-holiday periods? (ii) do movement rates of infected people change in different time periods?

We first fit the model to the data from 1944-1955 and 1956-1967 separately. As demonstrated in our simulated examples in Section 3.2 and 5.2, it is not possible to infer all the gravity model parameters at once. Hence, we set the parameters  $\tau_1$  and  $\tau_2$  equal to 1 and study the remaining key gravity model parameters  $\theta$  and  $\rho$ . The resulting 95% confidence regions for  $\theta'$  and  $\rho$  are provided in Figure 7. Based on this, we conclude that the change in parameter values is statistically insignificant for these two different time periods. Figure 8 shows confidence regions obtained by the GP emulator-based approach by fitting the model to the parts of the data corresponding to periods of holidays and non-holidays. As can be seen, the two regions are almost identical again, indicating that any change in the number of incidences for holidays and non-holidays is not due to the change in the way the infection spreads between cities of the metapopulation.

Since the matrix  $M = \{m_{kj}\}$ , where  $m_{kj} = \theta' N_{kt}^{\tau_1} \sum_{t=1}^T \frac{(I_{jt})^{\tau_2}}{d_{kj}^{\rho}}$  is interpreted as a matrix of the amount of movement, sum of  $k$ -th row of  $M$  represents the amount of infected individuals leaving city  $k$  while sum of  $k$ -th column is the number of infected people coming to city  $k$ . Using samples for  $\theta'$  and  $\rho$ , we easily obtain a sample for the spatial flux of infection for selected cities. In Table 1, we report our estimates with corresponding credible regions based on this analysis. We use the posterior median as point estimates. For example, we estimate the average number of emigrating infections during the holiday periods each week to be equal to 31.1 for London. Below the estimate, we report a 95% credible interval for it which is (4.4, 479.1). Based on these estimates, the mobility of the infection appears to be less during the periods of holidays. Table 2 shows estimates of the average amount of transit infections each week for years 1944-55 and 1956-67. We see here that the infection appears to move less

Table 1: Estimated amount of average movement in two weeks

City	From		To	
	Holiday	Non-Holiday	Holiday	Non-Holiday
London	31.1	46.6	34.4	49.8
	(4.4, 479.1)	(6.6, 744.7)	(4.6, 564.9)	(6.9, 823.9)
Birmingham	7.5	10.8	7.5	11.5
	(1.2, 72.9)	(1.8, 110.6)	(1.2, 74.7)	(1.9, 115.8)
Manchester	7.8	10.3	9.1	10.8
	(1.0, 151.4)	(1.4, 180.9)	(1.2, 162.9)	(1.5, 189.1)
Blackpool	0.8	1.1	0.6	0.7
	(0.1, 6.7)	(0.2, 8.8)	(0.1, 5.2)	(0.1, 6.1)

during the later years. Note that none of the differences are statistically significant. Note, however, that these analysis are not meant to represent a comprehensive epidemiological treatment, rather they are meant as an illustration of how the GP-approach permits rigorous statistical inference for the gravity TSIR model.

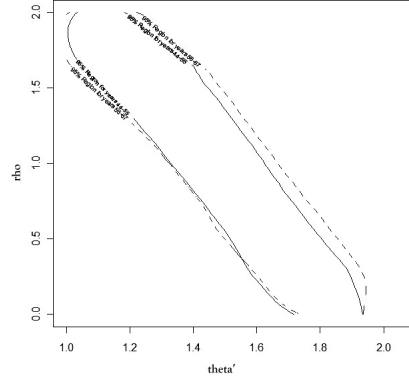


Figure 7: 95% C.I. for  $(\theta', \rho)$  obtained via fitting GP emulator to a part of the data: Solid line outlines the confidence region for parameters when data for years from 1944 -55 is used; Dashed line outlines the confidence region for parameters when data for years from 1956 -67 is used.

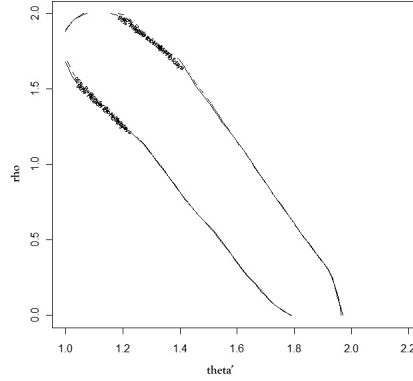


Figure 8: 95% C.I. for  $(\theta', \rho)$  obtained via fitting GP emulator to a part of the data: Solid line outlines the confidence region for parameters when data only holidays periods is used; Dashed line outlines the confidence region for parameters when data for only non-holiday periods is used.

Table 2: Estimated amount of average movement in two weeks

City	From		To	
	1944-55	1956-1967	1944-55	1956-1967
London	49.3	42.3	53.6	45.4
	(8.3, 532.4)	(5.4, 679.9)	(8.9, 587.5)	(5.6, 779.8)
Birmingham	11.6	10.3	12.8	10.4
	(2.1, 80.2)	(1.5, 109.7)	(2.3, 88.4)	(1.5, 109.5)
Manchester	11.0	9.9	12.9	9.7
	(1.8, 139.1)	(1.2, 189.2)	(2.1, 157.8)	(1.2, 184.6)
Blackpool	1.2	1.0	1.0	0.65
	( 0.2, 7.8)	(0.1, 9.4)	(0.2, 6.4)	( 0.1, 5.9)

## 7 Discussion

Complex models are very useful for representing physical phenomena, whether the phenomena is the spread of an infectious disease or the change in sea surface temperatures in the Atlantic. As is well known, it is not always possible for every aspect of such complicated phenomena to be modeled accurately; certain key characteristics of the process necessarily have to be focal points of the modeling effort. However, these key characteristics are not typically the focus of a statistical inferential procedure that uses a traditional likelihood-based approach. The approach we have developed in this paper addresses this point by providing a flexible inferential method that directly takes into account the characteristics of the process that are most important to scientists. Even though focusing on different summary statistics can lead to different estimates, parameter inference based on our approach produces an improved model fit to the biologically interesting features of the infectious disease dynamics. In addition to the flexibility this provides, we find that our approach is also computationally tractable in situations where traditional likelihood-based inference is not.

Computer model emulation and calibration is an active area of research (cf. Bayarri et al., 2007a; Bhat et al., 2010; Bhattacharya, 2007; Conti and O’Hagan, 2010; Higdon et al., 2008; Rougier and Beven, 2009; Rougier et al., 2009; Sansó and Forest, 2009) but most of this work has focused on deterministic models. In this paper, we describe an emulation-calibration approach for probabilistic models. In our view, therefore, our paper makes the following main contributions : (1) a general inferential approach that focuses on characteristics (summary statistics) of a process; (2) a method for statistical inference when the likelihood is intractable *and* simulation from the probability model is expensive; (3) a study of a particular model for measles dynamics, the gravity TSIR model, using the approach we have developed.

In the context of measles in the pre-vaccination era, our method allows us to study some interesting aspects of the dynamics based on the gravity TSIR model. For instance, we find that there does not appear to be a significant change in the gravity parameters for the school holiday periods versus non-holidays which means that we do not have enough evidence of a change in the dynamics of measles between these different periods. The gravity parameter estimates for years 1944-55 and 1956-67 were also not statistically different.

The methodology we have described in this paper is particularly useful in cases where simulation from a probability model might be too expensive to allow the use of other popular inferential approaches like ABC. It is worth noting that our approach works well when the parameter dimensionality is small, but is generally infeasible for parameter dimensions greater than around five to eight depending on the model complexity. Our approach is widely

applicable for inference in computationally expensive but biologically realistic models. In principle, whenever a likelihood is expensive to evaluate or when traditional Bayes approaches does not capture the most scientifically relevant features of the model, our method provides a way to incorporate important characteristics in a computationally tractable inferential approach.

## Acknowledgements

This work was supported in part by a grant from the Bill and Melinda Gates Foundation.

## Appendix A: Likelihood function for the gravity model

We provide below details of the likelihood function for the gravity model. For  $t \geq 2$ ,

$$\begin{aligned} L(I_{k(t+1)}, S_{k(t+1)}, L_{kt} | I_{kt}, S_{kt}) &= \\ &= L(I_{k(t+1)}, S_{k(t+1)} | I_{kt}, S_{kt}, L_{kt}) L(L_{kt} | I_{kt}, S_{kt}) = \\ &= L(I_{k(t+1)}, S_{k(t+1)} | I_{kt}, S_{kt}, L_{kt}) \frac{L_{kt}^{m_{kt}-1} \exp(-L_{kt})}{\Gamma(m_{kt})} \end{aligned}$$

where, as defined before in Equation 3,

$$m_{kt} = \theta N_{kt}^{\tau_1} \sum_{j=1, j \neq k}^K \frac{I_{jt}^{\tau_2}}{d_{kj}^{\rho}}.$$

$$\begin{aligned} L(I_{k(t+1)}, S_{k(t+1)} | I_{kt}, S_{kt}, L_{kt}) &= \\ &= L(S_{k(t+1)} | L_{k(t+1)}, I_{kt}, S_{kt}, L_{kt}) L(L_{k(t+1)} | I_{kt}, S_{kt}, L_{kt}) \\ &= L(S_{k(t+1)} | L_{k(t+1)}, S_{kt}) L(L_{k(t+1)} | I_{kt}, S_{kt}, L_{kt}) \\ &= L(S_{k(t+1)} | L_{k(t+1)}, S_{kt}) \frac{\lambda_{k,t+1}^{I_{k(t+1)}} \exp(-\lambda_{k,t+1})}{(I_{k(t+1)})!} \end{aligned}$$

where  $\lambda_{k,t+1} = \beta_t S_{kt} (I_{kt} + L_{kt})^\alpha$  as in Equation 1. Here,

$$L(S_{k(t+1)} | L_{k(t+1)}, I_{kt}) = \begin{cases} 1, & \text{if } S_{k(t+1)} = S_{kt} + B_{kt} - I_{k(t+1)} \\ 0, & \text{otherwise.} \end{cases}$$



Finally, the likelihood function of the model (conditional on the first observations) at the parameters  $\theta, \tau_1, \tau_2, \rho$  can be written as

$$\begin{aligned} L(\{I_{kt}\}_{k \geq 1, t \geq 2}, \{S_{kt}\}_{k \geq 1, t \geq 2} | \{I_{k1}\}_{k \geq 1}, \{S_{k1}\}_{k \geq 1}) = \\ = \int_S \int_L \prod_{t=2}^T \prod_{k=1}^K L(I_{k(t+1)}, S_{k(t+1)}, L_{kt} | I_{kt}, S_{kt}) dS dL. \end{aligned}$$

## Appendix B: Details of constructing an emulator

We first review our notation. Let  $Z = (z_1, \dots, z_K)$  denote the vector of summary statistics calculated using the observations. In our examples, these summary statistics are the proportions of zeros for each community and  $K$  is the total number of communities in our space-time data. We then let  $\Theta = (\theta, \tau_1, \tau_2, \rho)$  denote the vector of gravity parameters,  $Y(\Theta) = (Y_1(\Theta), \dots, Y_K(\Theta))$  denote the vector of summary statistics obtained using a simulation from the gravity model with the parameters  $\Theta$ . We choose a fixed grid  $\Omega = (\Theta_1, \dots, \Theta_p)$  on the parameter space. We recall that we used a uniform grid in the four dimensional cube  $[0, 2]^4$  with 20 values of each parameter. In our examples,  $p$ , size of the grid, was therefore equal to  $20^4$ . For each  $i$ , we can calculate  $D_i = d(Z, Y(\Theta_i))$ , where, as was introduced before, the function  $d(\cdot, \cdot)$  returns the Euclidean distance between the vectors. In the first stage of our approach, we model  $\mathbf{D} = (D_1, \dots, D_p)$  as a Gaussian process:

$$\mathbf{D} | \Omega, \beta_{\mathbf{G}}, \xi_{\mathbf{G}} \sim N(X\beta_{\mathbf{G}}, \Sigma(\xi_{\mathbf{G}})),$$

where  $\beta_{\mathbf{G}}$  and  $\xi_{\mathbf{G}}$  are the parameters of the Gaussian process. The design matrix  $X$  has a form,

$$X = \begin{pmatrix} 1 & \Theta_1^T \\ 1 & \Theta_2^T \\ \vdots & \vdots \\ 1 & \Theta_p^T \end{pmatrix} = \begin{pmatrix} 1 & \theta_1 & \tau_{11} & \tau_{21} & \rho_1 \\ 1 & \theta_2 & \tau_{12} & \tau_{22} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \theta_p & \tau_{1p} & \tau_{2p} & \rho_p \end{pmatrix}$$

Elements of the covariance matrix  $\Sigma(\xi_G)$  are given by

$$\begin{aligned} (\Sigma(\xi_G))_{ij} &= \text{cov}(D_i, D_j) = \\ &= \begin{cases} \sigma_G^2 \exp(-\phi_G^2 \|\Theta_i - \Theta_j\|^2), & \text{if } i \neq j \\ \sigma_G^2 + \tau_G^2, & \text{otherwise.} \end{cases} \end{aligned}$$

Note, therefore, that  $\beta_{\mathbf{G}} = (\beta_{0G}, \beta_{1G}, \dots, \beta_{4G})$  are the regression parameters and  $\xi_{\mathbf{G}} = (\sigma_G^2, \tau_G^2, \phi_G)$  specify the covariance function of the Gaussian process. Since  $(D_1, \dots, D_p)$

and  $(\Theta_1, \dots, \Theta_p)$  are known, we can use maximum likelihood inference to obtain estimates  $(\hat{\xi}_{\mathbf{G}}, \hat{\beta}_{\mathbf{G}})$ .

For any new  $D^*$  obtained at an unknown  $\Theta^*$ , following a standard “plug-in” approach,

$$\begin{pmatrix} D^* \\ \mathbf{D} \end{pmatrix} = \begin{pmatrix} D^* \\ D_1 \\ D_2 \\ \vdots \\ D_p \end{pmatrix} \sim N(X_e \hat{\beta}_{\mathbf{G}}, \Sigma_e(\hat{\xi}_{\mathbf{G}})), \quad (7)$$

where

$$X_e = \begin{pmatrix} 1 & \Theta^{*T} \\ X \end{pmatrix} = \begin{pmatrix} 1 & \theta^* & \tau_1^* & \tau_2^* & \rho^* \\ X \end{pmatrix},$$

and

$$\Sigma_e(\xi_G) = \begin{pmatrix} \sigma_G^2 + \tau_G^2 & \mathbf{v}^T \\ \mathbf{v} & \Sigma(\xi_G) \end{pmatrix},$$

with  $\mathbf{v} = (v_1, \dots, v_p)$ ,  $v_i = \text{cov}(D^*, D_i)$  for  $i = 1, \dots, p$ .

From Equation 7, using standard multivariate normal theory (cf. Anderson, 1984), conditional distribution of  $(D^* | \mathbf{D}, \Omega, \Theta^*)$  can easily be derived,

$$D^* | \mathbf{D}, \Omega, \Theta^* \sim N(\bar{\mu}, \bar{\Sigma}),$$

where  $\bar{\mu} = \hat{\beta}_{0G} + \theta^* \hat{\beta}_{1G} + \tau_1^* \hat{\beta}_{2G} + \tau_2^* \hat{\beta}_{3G} + \rho^* \hat{\beta}_{4G} + \mathbf{v}^T \Sigma(\hat{\xi}_{\mathbf{G}})^{-1} (\mathbf{D} - X \hat{\beta}_{\mathbf{G}})$  and  $\bar{\Sigma} = \hat{\sigma}_G^2 + \hat{\tau}_G^2 - \mathbf{v}^T \Sigma(\hat{\xi}_G)^{-1} \mathbf{v}$ . We denote this predictive distribution by  $\eta(D^*; \Theta^*)$  and use as an emulator for the gravity model.

## References

- Agarwal, D. and Gelfand, A. (2005). Slice sampling for simulation based fitting of spatial data models. *Statistics and Computing*, 15(1):61–69.
- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons.
- Bartlett, M. S. (1956). Deterministic and stochastic models for recurrent epidemics. In Neyman, J., editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 4*, pages 81–109. University of California Press.

- Bayarri, M., Berger, J., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R., Paulo, R., Sacks, J., and Walsh, D. (2007a). Computer model validation with functional output. *Annals of Statistics*, 35(5):1874–1906.
- Bayarri, M., Berger, J., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C., and Tu, J. (2007b). A framework for validation of computer models. *Technometrics*, 49(2):138–154.
- Beaumont, M., Zhang, W., and Balding, D. (2002). Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025.
- Bhat, K., Haran, M., Tonkonojenkova, R., and Keller, K. (2011). Inferring likelihoods and climate system characteristics from climate models and multiple tracers. Technical report, Pennsylvania State University, Department of Statistics.
- Bhat, K. s., Haran, M., and Goes, M. (2010). Computer model calibration with multivariate spatial output. In Chen, M.-H., Dey, D. K., Müller, P., Sun, D., and Ye, K., editors, *Frontiers of Statistical Decision Making and Bayesian Analysis: In Honor of James O. Berger*, pages 168–184. Springer-Verlag Inc.
- Bhattacharya, S. (2007). A simulation approach to Bayesian emulation of complex dynamic computer models. *Bayesian Analysis*, 2(4):783–816.
- Bjørnstad, O., Finkenstädt, B., and Grenfell, B. (2002). Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*, 72(2):169–184.
- Black, F. (1989). Measles. *Viral infections of humans: epidemiology and control*. New York: Plenum publishing, 3:451–65.
- Bobashev, G., Ellner, S., Nychka, D., and Grenfell, B. (2000). Reconstructing susceptible and recruitment dynamics from measles epidemic data. *Mathematical Population Studies*, 8(1):1–29.
- Clarkson, J. A. and Fine, P. M. (1985). The efficiency of measles and pertussis notification in England and Wales. *International Journal of Epidemiology*, 14(1):153.
- Cliff, A., Haggett, P., and Smallman-Raynor, M. (1993). *Measles: an historical geography of a major human viral disease from global expansion to local retreat, 1840-1990*. Blackwell, Oxford [England]; Cambridge, Mass., USA.

- Conti, S. and O’Hagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *Journal of Statistical Planning and Inference*, 140(3):640–651.
- Craig, P., Goldstein, M., Rougier, J., and Seheult, A. (2001). Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454):717–729.
- Dietz, K. (1976). The incidence of infectious diseases under the influence of seasonal fluctuations. *Lecture Notes in Biomathematics*, 11:1–15.
- Earn, D., Rohani, P., and Grenfell, B. (1998). Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society B: Biological Sciences*, 265(1390):7.
- Ellner, S., Bailey, B., Bobashev, G., Gallant, A., Grenfell, B., and Nychka, D. (1998). Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. *American Naturalist*, 151(5):425–440.
- Fine, P. and Clarkson, J. (1982). Measles in England and Wales: an analysis of factors underlying seasonal patterns. *International Journal of Epidemiology*, 11(1):5.
- Finkenstädt, B., Bjørnstad, O., and Grenfell, B. (2002). A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. *Biostatistics*, 3(4):493–510.
- Finkenstädt, B. and Grenfell, B. (2000). Time series modelling of childhood diseases: a dynamical systems approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(2):187–205.
- Finkenstädt, B., Keeling, M., and Grenfell, B. (1998). Patterns of density dependence in measles dynamics. *Proceedings of the Royal Society B: Biological Sciences*, 265(1398):753.
- Flegal, J., Haran, M., and Jones, G. (2008). Markov chain Monte Carlo: can we trust the third significant figure. *Statistical Science*, 23(2):250–260.
- Glass, K., Xia, Y., and Grenfell, B. (2003). Interpreting time-series analyses for continuous-time biological models—measles as a case study. *Journal of theoretical biology*, 223(1):19–25.
- Grenfell, B., Bjørnstad, O., and Finkenstädt, B. (2002). Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. *Ecological Monographs*, 72(2):185–202.

- Grenfell, B., Bjornstad, O., and Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723.
- Grenfell, B. and Harwood, J. (1997). (Meta) population dynamics of infectious diseases. *Trends in Ecology & Evolution*, 12(10):395–399.
- Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583.
- Jones, G., Haran, M., Caffo, B., and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101(476):1537–1547.
- Keeling, M. J., Bjørnstad, O. N., and Grenfell, B. T. (2004). *Metapopulation dynamics of infectious diseases.*, pages 415–445. Elsevier.
- Kennedy, M. and O’Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.
- Liu, F., Bayarri, M., and Berger, J. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119–150.
- Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15324.
- McLean, A. and Anderson, R. (1988). Measles in developing countries Part I. Epidemiological parameters and patterns. *Epidemiology and infection*, 100(01):111–133.
- Neal, R. (2003). Slice sampling. *Annals of Statistics*, 31(3):705–741.
- Pritchard, J., Seielstad, M., Perez-Lezaun, A., and Feldman, M. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rougier, J. and Beven, K. (2009). Formal bayes methods for model calibration with uncertainty. *Applied Uncertainty Analysis for Flood Risk Management*, eds. Beven, K. and Hall, J., Imperial College Press/World Scientific.

- Rougier, J., Guillas, S., Maute, A., and Richmond, A. D. (2009). Expert knowledge and multivariate emulation: the thermosphere ionosphere electrodynamics general circulation model (TIE-GCM). *Technometrics*, 51(4):414–424.
- Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical science*, 4(4):409–423.
- Sansó, B. and Forest, C. (2009). Statistical calibration of climate system properties. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(4):485–503.
- Schenzle, D. (1984). An age-structured model of pre-and post-vaccination measles transmission. *Mathematical Medicine and Biology*, 1(2):169.
- Swinton, H. and Grenfell, G. (1998). Persistence thresholds for phocine distemper virus infection in harbour seal *Phoca vitulina* metapopulations. *Journal of Animal Ecology*, 67(1):54–68.
- Xia, Y., Bjørnstad, O., and Grenfell, B. (2004). Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *American Naturalist*, 164(2):267–281.